

Privilege Ends at the Prompt Box

A federal judge ruled this year that a defendant's chats with a consumer AI tool were not privileged. The ruling is narrower than the viral clips suggest, but for any company that keeps breach findings, HR investigations or patient data behind counsel, the lesson is the same: privilege does not survive a prompt box you don't control.

Daniel Rodriguez • Founder & CEO, Infragil • October 2026

What actually happened

Bradley Heppner, a former executive charged with securities and wire fraud, used the consumer version of Claude to work through his legal situation before he was arrested. Some of what he typed in came from conversations with his own lawyers. He later sent the outputs to his defense team.

When federal agents searched his home in November 2025, they seized devices holding 31 of those AI documents. His lawyers claimed privilege. On February 10, 2026, Judge Jed Rakoff of the Southern District of New York ruled from the bench that they were not protected, and followed with a written opinion on February 17 (*United States v. Heppner*, No. 25-cr-00503).

The reasoning came down to three points:

- **The AI is not a lawyer.** Privilege protects communications between a client and counsel. A chatbot is neither.
- **The chats were not confidential.** The court leaned on the consumer privacy policy, which allows the provider to collect inputs and outputs, train on them, and disclose them to third parties including government regulators.
- **No lawyer directed the work.** Heppner used the tool on his own, so it was not attorney work product. Sending the results to counsel afterward did not cleanse them.

Forwarding a chatbot transcript to your lawyer doesn't make it privileged. It makes it evidence your lawyer has also seen.

The judge did leave a door open. Had counsel directed Heppner to use the tool, he wrote, it might have functioned like a professional agent working under the lawyer's privilege, the same theory that covers accountants and forensic firms hired by counsel.

What the ruling does not say

The version going around social media is that using ChatGPT or Claude destroys privilege over anything you put in. That overstates it. Heppner is one district court ruling on one set of facts, and it is not binding on courts elsewhere.

On the same day, a magistrate judge in Michigan reached the opposite result on work product. In *Warner v. Gilbarco*, a plaintiff representing herself had used ChatGPT to prepare her case. The court held those materials were protected, reasoning that AI programs are tools, not persons, and that work product is only waived by disclosure to an adversary.

Legal commentators have also pushed back on Heppner's reasoning. The Harvard Law Review argued the court asked the wrong question: not whether Heppner sought legal advice from the AI, but whether he used it to get advice from his attorney.

Three facts did most of the work in Heppner, and every one of them is a choice a company controls:

1. **The tier.** It was a consumer account governed by consumer terms that allow training and disclosure. Enterprise agreements with no-training and confidentiality commitments are a different fact pattern, and no court has ruled on them yet.
2. **The direction.** No lawyer asked for the work.
3. **The timing.** The material was created first and handed to counsel later.

I run an AI-native company and use these tools every day. This is not an argument against AI. It is an argument for knowing which account, which terms and whose direction your sensitive work runs under.

Breach response: privilege was already fragile

Every CISO who has run a major incident knows the playbook. Call outside counsel first, have counsel hire the forensic firm, and keep the findings under privilege until you know what you are dealing with. Courts have been punching holes in that playbook for years.

In 2020, a federal court ordered Capital One to hand its Mandiant breach report to plaintiffs. Counsel had directed the work, but Mandiant was on a pre-existing business retainer and the report went to four regulators. In 2021, *Wengui v. Clark Hill* went the same way: the report was shared with leadership, IT and the FBI, and it contained remediation advice rather than legal advice.

Now add AI to the war room. An incident lead pastes log excerpts into a personal chatbot to summarize them. Someone drops counsel's memo into a consumer tool to turn it into board slides. The CISO drafts the regulator notice in an assistant on their phone at 2 a.m.

Under Heppner's reasoning, each of those can be a disclosure to a third party. Some commentators warn the waiver could reach the whole subject matter, not just the pasted text. And the data now sits on a vendor's servers, where it can be subpoenaed. Courts are already treating AI logs as evidence: the same district ordered OpenAI to produce 20 million de-identified ChatGPT logs in copyright litigation.

The forensic report was always the obvious target. The prompts your team typed while writing it are the new one.

HIPAA: a different law, the same failure

Privilege is a litigation problem. HIPAA is a compliance problem that shows up before anyone sues. A covered entity can only share protected health information with a vendor that has signed a Business Associate Agreement.

Consumer AI tiers don't offer one. OpenAI will not sign a BAA for ChatGPT Free, Plus, Team or self-serve Business, and its consumer ChatGPT Health product runs on consumer terms despite the name. The major vendors, Anthropic included, reserve BAAs for sales-managed enterprise and API agreements.

That means a nurse pasting a patient note into a personal chatbot, or a billing analyst asking one to clean up a claims export, is making an impermissible disclosure no matter how careful they are.

Healthcare also stacks the two problems. A hospital investigating a breach of patient data, under counsel, whose team pastes affected records into a consumer tool now has a HIPAA disclosure and a privilege question in the same keystroke.

The workplace exposure nobody inventoried

Companies route a lot through counsel precisely because they don't want it surfacing later. Breach root causes. Harassment and whistleblower investigations. Layoff planning. M&A diligence. Draft board minutes about a regulatory problem.

The people doing that work are also the people most likely to reach for AI to go faster. If they use a personal account, the company has a record it doesn't know exists, outside its retention schedule, outside its legal holds, and governed by terms it never negotiated. The first time anyone learns about it may be a discovery request asking for AI prompts and outputs.

Courts are already adjusting. Before Heppner was decided, the judge in the Warner case modified a protective order to bar uploading confidential documents to any AI platform. Expect that language to become standard.

This is the same pattern we lived through with cloud fifteen years ago. Shadow IT wasn't a technology problem. It was a visibility problem, and it got solved by giving people a sanctioned tool that was easier than the unsanctioned one.

The playbook

- 1. Inventory where AI touches sensitive work.** Legal, incident response, HR, compliance and anything with PHI, including personal accounts on personal devices.
- 2. Route privileged and regulated work to enterprise tiers only.** Negotiated terms with no training on your data, confidentiality obligations, defined retention, and a BAA wherever PHI is involved.
- 3. Put counsel in the driver's seat.** When AI is used in an investigation, counsel should direct it in writing, the same way they engage a forensic firm.
- 4. Run two tracks in breach response.** Keep the privileged legal investigation separate from business remediation, and decide in advance which AI tools each track may use.

5. **Bring AI logs into records management.** Prompts and outputs belong in your retention schedule and your legal hold process.
6. **Ask your AI vendors the hard questions.** Add training use, retention, government disclosure terms and BAA availability to your third-party risk reviews.
7. **Give people one plain rule.** If you wouldn't email it to a stranger, don't paste it into a personal chatbot.
8. **Brief your executives directly.** Anyone facing an investigation or lawsuit should talk to counsel before they talk to an AI.

Where Infragil fits

Infragil helps leadership teams adopt AI without handing away the protections they've spent years building: mapping where AI already touches privileged and regulated work, setting the tiers and rules, and treating AI providers as the third parties they are. That last piece is exactly what Skopos was built to assess. Learn more at infragil.com.

Daniel Rodriguez is Founder & CEO of Infragil and creator of Skopos, with 29 years in cybersecurity including CISO and VP roles. This article is not legal advice; talk to your counsel about your own situation.

Sources

- O'Melveny & Myers, [S.D.N.Y. First-of-its-Kind Ruling: AI-Generated Documents Are Not Privileged](#), February 2026
- Proskauer Rose, [SDNY Addresses Privilege and Work Product Implications of Using Unsecured Public AI Tools](#), February 2026
- Chapman and Cutler, [Federal Court Rules That AI-Generated Documents Are Not Protected by Privilege](#), February 2026
- Maynard Nexsen, [United States v. Heppner and AI Discovery](#), February 2026
- Proskauer Rose, [Michigan Federal Court Protects AI-Assisted Litigation Work Product](#), February 2026
- Harvard Law Review, [United States v. Heppner](#), March 2026
- Kirkland & Ellis, [Two Federal Courts Chart Diverging Paths on the Discoverability of LLM Interactions](#), March 2026
- Paul, Weiss, [Federal Courts Reach Different Outcomes on AI-Generated Materials](#), March 2026
- Lawfare, [AI and Privilege After United States v. Heppner](#), 2026
- Hunton Andrews Kurth, [Forensic Report Deemed Not Privileged: Capital One Ordered to Release Report](#), 2020
- Barnes & Thornburg, [Litigation Lessons From Cyberattacks \(Wengui v. Clark Hill, D.D.C. 2021\)](#), 2023
- HIPAA Journal, [Is ChatGPT HIPAA Compliant?](#), January 2026